

POSTPRINT

Grammatische Angaben in *elexiko* und ihre Modellierung

Annette Klosa und Carolin Müller-Spitzer

The project *elexiko* compiles an extensive, monolingual dictionary of contemporary German. This contribution deals with the grammatical data in this dictionary; it is not only described how these are arranged content-wise depending on corpus data, but also how they were modelled.

Das Projekt *elexiko* erarbeitet ein umfangreiches, einsprachiges Wörterbuch des Gegenwartsdeutschen. In diesem Beitrag geht es um die grammatischen Angaben in diesem Wörterbuch; es wird nicht nur erläutert, wie diese inhaltlich in Abhängigkeit vom Prinzip der Korpusbasiertheit gestaltet sind, sondern auch, wie sie modelliert wurden.

In diesem Beitrag soll es um die grammatischen Angaben und Prinzipien ihrer Modellierung im Projekt *elexiko* des Instituts für Deutsche Sprache (IDS) in Mannheim gehen. Zugleich erörtert diese Darstellung Chancen und Probleme des zu Grunde liegenden lexikografischen Prinzips der Korpusbasiertheit. Das Projekt *elexiko* erarbeitet ein einsprachiges Wörterbuch des Gegenwartsdeutschen von perspektivisch großem Umfang. Dieses soll im Internet publiziert werden, wobei schon in der Lemmaliste online recherchiert werden kann, die 300.000 Einträge umfasst (www.elexiko.de). Außerdem sind rund 700 ausgearbeitete Wortartikel online. Die weitere Erarbeitung und Publikation erfolgt zukünftig innerhalb von Arbeitsmodulen.

1 Grammatische Angaben in *elexiko*

Im Bereich der grammatischen Angaben eines *elexiko*-Artikels wird die Grammatik des Stichwortes immer in seinen einzelnen Lesarten beschrieben. Die grammatischen Angaben sind dabei so konzipiert, dass der Lexikograf zunächst die Wortart des untersuchten Stichwortes bestimmen muss. In Abhängigkeit von dieser

Entscheidung sind dann einzelne Angaben zu machen. Diese sind im Vergleich zur traditionellen Lexikografie umfangreicher. So werden alle vorgesehenen grammatischen Informationen zu einem Stichwort immer angegeben, egal ob sie erwartbar und vom Nutzer selbst bildbar wären, weil sie z. B. regelmäßig sind, oder umgekehrt. Damit kommt das Projekt einer Forderung der Metalexikografie nach, die u. a. Wiegand (1985: 96) formuliert hat, nämlich "dass zur Gewinnung einer grammatischen Information möglichst wenig Zweitnachschrägegänge nötig werden" (vgl. auch Mugdan 1989: 519).

1.1 Chancen des Prinzips der Korpusbasiertheit

Eines der zentralen Prinzipien bei der lexikografischen Arbeit an *lexiko* ist das Prinzip der Korpusbasiertheit. Zentraler Ansatz für das Projekt ist, alle Angaben, egal ob zur Grammatik, zur Bedeutung, zur Verwendung usw., konsequent aus dem *lexiko*-Textkorpus zu gewinnen. Dies hat seinen Grund darin, dass sich das Projekt für den aktuellen und öffentlichen Sprachgebrauch interessiert, d. h. dafür, wie die Sprecher und Sprecherinnen der deutschen Sprachgemeinschaft üblicherweise miteinander kommunizieren.

Das *lexiko*-Korpus ist ein dynamisches, also ein so genanntes Monitorkorpus, welches regelmäßig erweitert und aktualisiert wird, um die jeweils neuesten Entwicklungen verfolgen und damit aktuelle Beschreibungen liefern zu können. Für die Analyse des öffentlichen Sprachgebrauchs bietet sich die Auswertung von Zeitungstexten besonders an, da diese sprachlich den Fokus der öffentlichen Diskussionen widerspiegeln. Das *lexiko*-Korpus besteht daher hauptsächlich aus Zeitungs- und Zeitschriftentexten. Diese Textsorte setzt sich nicht nur aus einer Vielzahl an verschiedenen Texttypen (Reiseberichte, Leserbriefe, Anzeigen, Auszüge aus Romanen, Amtstexte usw.) zusammen, sondern deckt darüber hinaus auch zahlreiche Themen aus Politik, Kultur, Wissenschaft, Sport, Tourismus u. a. ab und ist somit als Datengrundlage für die Zielsetzung von *lexiko* besonders gut geeignet.

Das *lexiko*-Korpus ist Teil des IDS-Korpus, das weltweit größte Textkorpus zur deutschen Schriftsprache mit bald 2 Milliarden laufenden Textwörtern. Das Korpus wird mithilfe verschiedener intelligenter automatischer Recherche- und Analysewerkzeuge, der so genannten COSMAS-Plattform, erschlossen. Dabei handelt es sich um ein komplexes Werkzeug zur Strukturierung von Massendaten, das zu vielen verschiedenen Zwecken eingesetzt werden kann, z. B. für einfache Wortsuchen oder Volltextsuchen, aber auch für chronologische Sortierungen oder zum Erkennen von Gesetzmäßigkeiten bei Wortzusammensetzungen und vielem mehr.

lexiko Lesartenansicht - tipptopp - Microsoft Internet Explorer

Funktion(en) im Satz: attributiv

Kommentar(e)

Kommentar:
Im lexiko-Korpus ist **tipptopp** selten attributiv belegt:
Sie mögen den kleinen Spanier. In seiner Freizeit trinke er manchmal ein Bier zuviel, sagen sie, aber er sei ein "tipptopper" - Arbeiter und habe überdes ein gutes Herz. 03.08.1998, S. 13, (Zürcher Tagesanzeiger, Bellevue, 20.23 Uhr (Serie).)

Schließen

adverbial

Kommentar(e)

Kommentar:
Im lexiko-Korpus wird **tipptopp** fast immer adverbial gebraucht:
Wenn die SED-Bonzen in ihren Limousinen nach Pankow rauschten, sahen sie im Vorbeifahren nur eine Seite der Nebenstraßen und "dachten, es sieht alles **tipptopp** aus".
(die tageszeitung, 01.03.1990, S. 28, Gründergeist am Prenzlauer Berg.)

Schließen

prädikativ

Kommentar(e)

Kommentar:
Im lexiko-Korpus ist **tipptopp** selten in prädikativer Verwendung belegt:

Abbildung 1. Artikel "tipptopp" in *lexiko*

Aus dem *lexiko*-Textkorpus sind nicht nur frequenzbasierte Angaben dazu zu gewinnen, welche Wörter überhaupt in die Lemmaliste aufgenommen werden, sondern es werden natürlich auch Informationen zu einem Suchwort, die in den entsprechenden Artikeln in schon publizierten Wörterbüchern noch nicht enthalten sind, ermittelt. Daneben gibt es Fälle, in denen der Korpusbefund eine Präzisierung der bereits in gedruckten Wörterbüchern enthaltenen Informationen ermöglicht. Innerhalb der grammatischen Beschreibung sind beispielsweise eine Reihe von Verwendungshäufigkeitsangaben vorgesehen, die nur mithilfe von Korpusrecherchen erarbeitet werden können. Als Beispiel sollen hier die Funktionen des Adjektivs *tipptopp* im Satz untersucht werden. Im *lexiko*-Korpus sind 26 Belege für attributives und flektiertes *tipptopp* enthalten und 376 Belege für unflektiertes und hauptsächlich adverbial gebrauchtes *tipptopp*. Im Wortartikel zu *tipptopp* wird dieses auffällige Verhältnis von etwa 1:14 zwischen attributiver und (überwiegend) adverbialer Verwendung unter den grammatischen Angaben deshalb in Form von entsprechenden Kommentaren gesondert vermerkt. Dabei werden die unterschiedlichen Verwendungshäufigkeiten mithilfe eines so genannten "standardisierten Häufigkeitswortes" nach einem Vorschlag von Bergenholtz und Mugdan (1984b: 63) markiert: "fast immer" für 90 % der Belege und mehr, "meist" für 70 % bis unter 90 % der Belege, "oft" für 50 % bis unter 70 % der Belege, "auch" für 30 % bis unter 50 % der Belege, "manchmal" für 10 % bis unter 30 % der Belege, "selten" für unter 10 % der Belege, "genauso" für genau 50 % der Belege. Entsprechend lauten die Einträge in der Online-Fassung des Artikels (vgl. Abbildung 1).

Die *lexiko*-Artikelstruktur ermöglicht es schließlich, im Bereich der Grammatik Angaben, die in traditionellen Wörterbüchern durchaus enthalten, jedoch nicht recherchierbar sind, gezielt zugreifbar zu machen. Systematisch wird beispielsweise im *lexiko*-Artikel nicht nur die Wortart angegeben, sondern es werden auch das syntaktische Verhalten bestimmter Wortarten, Valenzeigenschaften bestimmter Wortarten oder morphologische Charakteristika erfasst.

Die *lexiko*-Artikelstruktur ist in einer XML-Dokumenttypdefinition (kurz: DTD) festgehalten. Durch die Anwendung der DTD werden die Daten inhaltlich ausgezeichnet (vgl. hierzu Abschnitt 2. zur Modellierung der grammatischen Angaben). Die DTD wurde z. T. schon anhand von Musterartikeln entwickelt, damit sie den vorgesehenen lexikografischen Inhalten wirklich angemessen ist. Im Bereich der Grammatik basieren die in der DTD verankerten einzelnen Angaben daneben grundsätzlich auf (Zifonun et al. 1997), einer dreibändigen "Grammatik der deutschen Sprache", die am Institut für Deutsche Sprache erarbeitet wurde. Die *lexiko*-DTD versucht in diesem Bereich, die in dieser Grammatik enthaltenen Kategorisierungen und Beschreibungen so explizit wie möglich zu machen. So wird etwa für alle Adjektive in den Wortartikeln jeweils auf die einzelne Lesart bezogen vermerkt, ob sie steigerbar sind oder nicht, wie das Beispiel *international* zeigt (vgl. Abbildung 2).

elexiko Lesartenansicht international - Microsoft Internet Explorer

international

Lesart 'überstaatlich'

Lesartenübergreifende Angaben:
 Lesart 'überstaatlich'
 Lesart 'zwischenstaatlich'

Bedeutungs- erläuterung	Semantische Umgebung u. lexikalische Mitspieler	Typische Verwendungen	Sinnverwandte Wörter	Besonderheiten des Gebrauchs	Grammatik
Grammatik <small>Info</small>					
Wortart:	Adjektiv (deklinierbar)				<small>Beleg(e)</small>
Komparativ: Superlativ:	internationaler (am) internationalsten				<small>Beleg(e)</small>
Funktion(en) im Satz:	attributiv adverbial prädikativ				<small>Beleg(e)</small>

Druckversion
des gesamten
Artikels

Abbildung 2. Artikel "international", Lesart 'überstaatlich' in *elexiko*

Der Rückbezug auf das Grammatikmodell, das der grammatischen Beschreibung im Projekt *lexiko* zu Grunde liegt, ist dadurch gewährleistet, dass die Online-nutzer immer wieder auf die Möglichkeit hingewiesen werden, in Grammis, dem Grammatischen Informationssystem des Instituts für Deutsche Sprache, das ebenfalls über das Internet recherchierbar ist (<http://hypermedia.ids-mannheim.de/grammis>), nachzuschlagen. So beziehen sich unsere grammatischen Angaben explizit auf ein bestimmtes Grammatikmodell mit bestimmten Beschreibungskategorien (vgl. Bergenholtz/Mugdan 1984a: 16).

1.2 Probleme, die sich aus der Korpusbasiertheit ergeben

Neben den Vorteilen, die eine korpusbasierte Erarbeitung der grammatischen Angaben bietet, entstehen natürlich auch Probleme, wenn man ausschließlich mit Korpusdaten arbeitet. So muss auch das Projekt *lexiko* damit umgehen, dass sich Empirie, Norm und lexikografische Ansprüche nicht immer decken. Besonders zwei Konfliktgebiete (vgl. Farø 2005) haben sich dabei während der Erarbeitung der Wortartikel für *lexiko* als problematisch erwiesen: a) Das Korpus bietet einen Befund, welcher die aus der traditionellen Grammatik in die grundsätzliche Artikelstruktur bzw. DTD übernommene Kategorisierung oder die vorgesehenen Informationen nicht bestätigt; b) Das Korpus hat Lücken.

Als Beispiel für das Phänomen, dass Korpusbefund und angenommene grammatische Kategorisierungen nicht übereinstimmen, sollen im Folgenden bestimmte flexivische Eigenschaften der Adjektive dienen. So wird traditionell, und auch in Zifonun et al. (1997), zwischen deklinablen und indeklinablen Adjektiven unterschieden, wobei als indeklinabel gemeinhin Farbadjektive wie *rosa* gelten. Diese Unterscheidung soll nach der *lexiko*-Artikelkonzeption für jedes beschriebene Adjektiv getroffen und im Artikel festgehalten werden. Dies hat den Zweck, dass der Gesamtbestand an Artikeln später einmal nach allen deklinablen oder allen indeklinablen Adjektiven durchsucht werden kann.

Für das Adjektiv *rosa* ergibt die Recherche im *lexiko*-Korpus fast 5.000 Treffer für unflektiertes Vorkommen in attributiver Stellung (*rosa Rosen*) und knapp 20 Treffer für flektiertes Vorkommen in der gleichen Stellung (*rosane Rosen*). Das Adjektiv *rosa* ist damit nicht wirklich indeklinabel; im Artikel wird also eingetragen "deklinabel". Darüber hinaus gibt es aber zusätzlich einen Hinweis darauf, dass *rosa* gemeinhin als indeklinabel gilt, und es sich hier also wohl um ein Normproblem handelt (vgl. Abbildung 3).

lexikon-lesartenansicht-rosa - Microsoft Internet Explorer

rosa Lesart 'blassrot'

Lesartenübergreifende Angaben:
Lesart 'blassrot'
 Lesart 'schneef'
 Lesart 'bodematisch'

Nahezu alle Verteilung:
 Diese Lesart ist in Texten des folgenden Sprachraumes belegt:
 BRD, Österreich, Schweiz

Bedeutungs- erläuterung	Semantische Umgebung u. lexikalische Mitspieler	Typische Verwendungen	Sinverwandte Wörter	Erscheinungen des Gebrauchs	Grammatik
<p>Grammatik Info</p> <p>Wortart: Adjektiv (deklinierbar)</p> <p>nicht steigerbar</p>					
<p>Hinweis(e)</p> <p>Sprachreflexiver Hinweis: Das Adjektiv rosa gilt als indeklinabel. Wird es dekliniert, wird dies teilweise als Normproblem eingestuft.</p> <p>Weitere Informationen: (Duden - Richtiges und gutes Deutsch, Wörterbuch der sprachlichen Zweifelsfälle, 5., neu bearbeitete Auflagen, Herausgegeben von der Dudenredaktion, Dudenverlag, Mannheim/Leipzig/Wien/Zürich 2001, (Duden Band 9), S. 729.) (Wahrig Fehlerfreies und gutes Deutsch, Das zuverlässige Nachschlagewerk zur Klärung sprachlicher Zweifelsfälle, Von Jürgen Dittmann, Rolf Thieroff und Ulrich Adolphs, Bertelsmann Lexikon Institut, Gütersloh/München 2003, S. 371 und 378.)</p> <p>Schließen</p>					

Abbildung 3. Artikel "rosa", Lesart 'blassrot' in *lexiko*

Werden durch die Korpusanalyse Abweichungen zu den angenommenen Kategorien oder Eigenschaften festgestellt, müssen sie in *lexiko* also erfasst werden. Dies geschieht sowohl durch eine insgesamt umfänglichere und ausführlichere Beschreibung wie auch durch die Möglichkeit, durch ergänzende Belege, Hinweise und Kommentare auf den abweichenden Korpusbefund hinzuweisen. Mit dieser korpusbasierten Beschreibung ist gegenüber der traditionellen Lexikografie der Fortschritt zu erzielen, dass traditionelle Kategorisierungen anhand eines Korpus hinterfragt werden und der tatsächliche Sprachgebrauch (im *lexiko*-Korpus) dokumentiert wird. Das Projekt *lexiko* stellt so im Idealfall im Lauf der Bearbeitungs- und Publikationszeit Material bereit, das Grammatiker später einmal benutzen können, wenn sie eine wirklich synchrone, "empirische" Grammatikschreibung anstreben. Auf der anderen Seite kann nicht innerhalb der lexikografischen Beschreibung von *lexiko* die Grammatik komplett neu geschrieben werden – im Einzelfall muss man also, nicht zuletzt aus Gründen der Praktikabilität, sicherlich abwägen, welche Befunde man im Artikel festhalten will und welche nicht, denn Massendaten so detailliert auszuwerten, wie hier vorgeführt, kostet viel Zeit.

Ein weiteres Problem für die Erfassung der grammatischen Angaben in *lexiko* ist vor folgendem Hintergrund zu sehen:

When you analyze corpus data, you constantly find that inflectional forms given in the dictionary are in fact not attested in actual text productions, or the opposite situation of inflectional forms occurring that are not part of the official paradigm. The problem is primarily one of interpretation: is the absence of an inflectional form an indication that it does not exist, or is it an indication that the corpus is simply not large enough? (Trap-Jensen 2002)

Wie geht *lexiko* damit um, wenn das zu beschreibende grammatische Phänomen oder die gesuchte Flexionsform zu einem Stichwort gar nicht belegt ist, das *lexiko*-Korpus also Lücken hat?

Zu starken Verben werden etwa traditionellerweise in Wörterbuchartikeln bestimmte finite Formen, die Stammformen, angegeben, und das ist auch in *lexiko* so. Als Beispiel soll nun das Verb *weben* dienen; lässt man es mithilfe von COSMAS im *lexiko*-Korpus suchen, generiert der Wortformenoperator folgende Liste an Zeichenfolgen, die der Grundform "weben" zugeordnet werden (vgl. Abbildung 4).

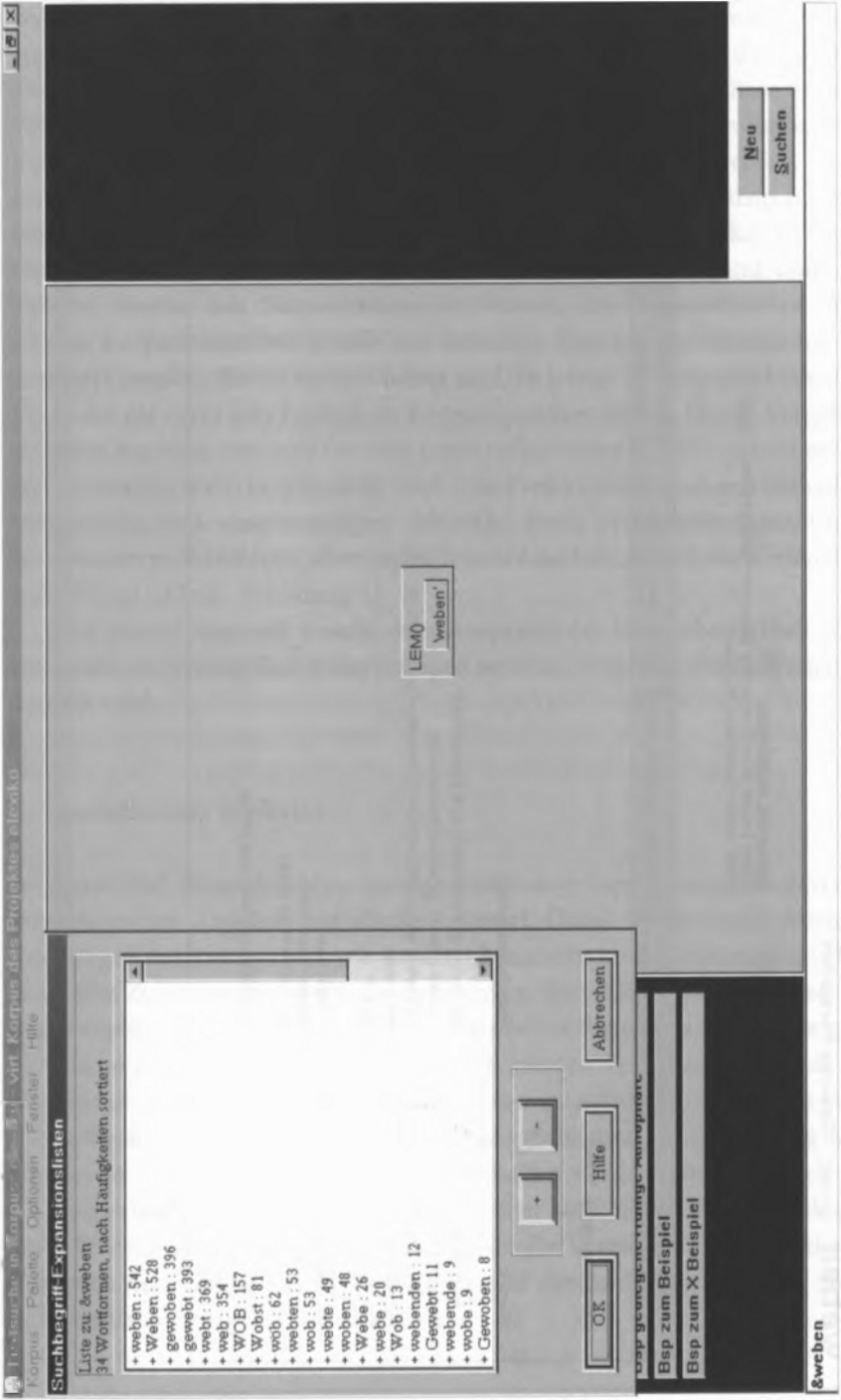


Abbildung 4. Ergebnissenster Wortformensuche zu “&weben” in COSMAS

lexiko Lesartenansicht: weben Microsoft Internet Explorer

weben

Lesarten 'Gewebe herstellen'

Lesartenüberarbeitende Anbieter:
Lesart 'Gewebe herstellen':
Lesart 'gestalten'

Nationale Verteilung:
Diese Lesart ist in Texten des folgenden Sprachraumes belegt:
BRD, DDR, Österreich, Schweiz

Bedeutungs- erläuterung	Semantische Umgebung u. lexikalische Mitspieler	Typische Verwendungen	Sinnverwandte Wörter	Besonderheiten des Gebrauchs	Grammatik
<p>Grammatik Info</p> <p>Wortart: Verb (schwankend)</p> <p>3. Pers. Sing. Ind. Prät. Aktiv: Partizip Perfekt: 3. Pers. Sing. Konj. Prät. Aktiv: Perfektbildung: Passiv:</p> <p><i>(er/sie/es) webte (meist), web (manchmal) gewebt (meist), gewoben (manchmal) (er/sie/es) webte, wöbe (nicht im <i>lexiko</i>-Korpus belegt) mit haben bildbar</i></p> <p>Satzbaupläne: jemand webt jemand webt etwas jemand webt <i>an</i> etwas jemand webt etwas <i>aus</i> etwas</p> <p>Hinweise(s)</p> <p>Belege(s)</p> <p>Belege(s)</p> <p>Druckversion des gesamten Artikels</p>					

Abbildung 5. Artikel "weben", Lesart 'Gewebe herstellen' in *lexiko*

Eine Überprüfung aller Belege zur Zeichenfolge *wöb* ergibt beispielsweise, dass es sich hierbei nicht um die Verbform der 3. Person Singular Konjunktiv Präteritum zu *weben* handelt (verkürzt um das auslautende *-e*), sondern um das Kürzel eines Journalisten der Berliner Zeitung. Somit ist im *lexiko*-Korpus diese finite Form des Flexionsparadigmas von *weben* nicht belegt. Die Struktur von Verbartikeln in *lexiko* sieht aber vor, dass mehrere finite Formen verzeichnet werden, und zwar immer die 3. Person Singular Indikativ Präteritum und das Partizip II; daneben die 3. Person Singular Konjunktiv Präteritum, falls der Stammvokal im Konjunktiv Präteritum nicht mit dem im Indikativ Präteritum übereinstimmt, und schließlich bei Verben mit Stammwechsel im Präsens die entsprechenden Formen. Strenge Korpusbasiertheit würde nun bedeuten, dass nur die Formen im Artikel vermerkt werden, die im Korpus belegt sind. In letzter Konsequenz bedeutet das, dass man gar nicht von Lücken im Korpus sprechen dürfte. Dieses Vorgehen hat aber den Nachteil, dass eine für viele potenzielle Nutzer des Wörterbuches wichtige Information nicht bereitgestellt wird. Das Projekt *lexiko* hat sich deshalb dafür entschieden, u. a. eine sozusagen "fehlende" Form in Verbparadigmen kompetenzgestützt nachzutragen, aber zugleich zu vermerken, dass sie im *lexiko*-Korpus nicht belegt ist (vgl. Abbildung 5).

Auf diese Weise soll sowohl dem Anspruch der Korpusbasiertheit wie den Interessen der Nutzer Rechnung getragen werden, ohne dass *lexiko* zugleich präskriptiv wird.

2 Modellierung in *lexiko*

Im ersten Teil dieses Aufsatzes wurden bestimmte Aspekte aus dem Bereich der grammatischen Angaben aus *lexiko* erläutert. Dabei wurden auch Ausschnitte aus fertigen Artikeln gezeigt, wie sie allen Benutzern im Internet zugänglich sind. In diesem Zusammenhang war die Rede davon, dass alle Angaben in unseren Artikeln explizit markiert sind und den potenziellen Benutzern so für eine gezielte Recherche zur Verfügung stehen sollen. Im zweiten Teil dieses Aufsatzes soll es nun darum gehen, den Projektbereich genauer zu erläutern, in dem die wesentlichen Voraussetzungen für diese späteren Möglichkeiten der Darstellung und der Recherche der lexikografischen Daten geschaffen werden. Dieser Bereich ist für die Wörterbuchbenutzer nicht sichtbar, ist aber wesentlich für die Lexikografen beim Erarbeiten der Artikel und schafft eben die Voraussetzungen für die Möglichkeiten der Publikation der Artikel. Es geht demnach hier mehr um die texttechnologischen Aspekte von *lexiko*.

2.1 Anforderungen an die Modellierung lexikografischer Daten

Jedes Projekt, bei dem es um die Erfassung und Verarbeitung von Textdaten geht, steht heutzutage vor Fragen wie:

- In welcher Weise sollen die Textdaten erfasst werden?
- In welcher Form werden die Textdaten gespeichert?
- Mit welchen Hilfsmitteln wird das Textmaterial für die weitere Arbeit oder Veröffentlichung erschlossen? (Vgl. Büchel/Schröder 2001: 8)

In lexikografischen Projekten ist der dritte Punkt – der Zugriff auf die Daten – besonders zentral, da die Realisierung der Zugriffsmöglichkeiten ein wichtiges Unterscheidungskriterium für gedruckte vs. elektronische Wörterbücher darstellt. "From a user's perspective, the most innovative aspect of EDs is probably the retrieval system" (Schryver 2003: 146). Dabei geht es nicht allein um die Formulierung von bestimmten Suchanfragen, sondern auch um die flexible Präsentation des Suchergebnisses. Daher wird gefordert, abhängig von Benutzungssituationen aus einem Wörterbuchartikel nur die gesuchten Angaben anzeigen zu können. "It is one thing to be able to store ever more data, but another thing entirely to present just the data users want to response to a particular look-up." (Schryver 2003: 178, ähnlich Klosa 2001: 97) "Die Benutzerschnittstelle von digitalen Wörterbüchern" sollte daher "an Typen von Benutzungssituationen adaptierbar sein." (Storrer 2001: 64) Die Voraussetzungen für diese flexiblen Zugriffsmöglichkeiten müssen bei der Modellierung geschaffen werden. Denn die Art der Modellierung und der Datenaufbereitung entscheidet darüber, wie gezielt und genau auf diese Daten später zugegriffen werden kann. "What you mark is what you get" (Gloning/Welter 2001: 128), so formulieren es Gloning/Welter bezogen auf eine XML-basierte Modellierung. Neben den gezielten Zugriffsmöglichkeiten sind folgende Anforderungen an die Modellierung lexikografischer Daten noch zusammenfassend zu nennen: Daten aus unterschiedlichen Medien müssen in eine Datenbasis integriert werden können und lexikografische Daten müssen an externe Ressourcen wie z. B. Textkorpora angebunden werden können.

Die bis jetzt genannten Anforderungen an ein Modellierungskonzept beziehen sich vor allem auf die Funktionalität elektronischer Wörterbücher. Ein weiterer wichtiger Bereich ist die Computerunterstützung im lexikografischen Prozess. Auch für die Lexikografen ist dabei ein gezielter Zugriff auf die lexikografischen Daten wichtig, um z. B. leichter und verlässlicher Konsistenzprüfungen der Beschreibungssprache einzelner Angaben vornehmen zu können. Daneben sollten die Lexikografen in der formalen Einhaltung der Artikelstruktur unterstützt werden. So fordert Wiegand, dass die Verantwortung für die Einhaltung des Artikelformats an ein Computerprogramm abgetreten werden könne, da dies bei

anspruchsvollen Wörterbüchern eine Entlastung für den Lexikografen bedeute (Wiegand 1998: 232). Traditionell werden Schreibanweisungen für die Artikelarbeit in Instruktionsbüchern oder Manualen festgehalten. Das Problem in der lexikografischen Praxis ist jedoch oft, dass besonders eine komplexe Artikelstruktur ohne umfangreiche technische Unterstützung kaum einzuhalten ist.

Konventionen, die die standardisierten Wörterbuchartikel erfüllen müssen, werden normalerweise in Instruktionsbüchern festgehalten. Das Einhalten dieser Konventionen wird manuell und oft nur stichprobenartig überprüft. In Anbetracht der hochkomplexen Textform 'Wörterbuchartikel' ist eine redaktionelle Betreuung eines Wörterbuchs mit mehr als 60000 Artikeln allerdings sehr schwierig und letztendlich sind inhaltliche und formale Mängel kaum auszuschließen. (Heyn 1992: 187)

Außerdem sollten lexikografische Angaben ohne Strukturanzeiger, d. h. ohne Kommata, typografische Auszeichnungen etc. eingegeben werden können. Diese Unterstützungen können eine wesentliche Erleichterung für die lexikografische Arbeit darstellen.

Im Zusammenhang mit der Modellierung lexikografischer Daten ist auch eine mögliche Schnittstelle zur automatischen Sprachverarbeitung zu beachten. Denn das Zusammenwirken von Lexikografie und automatischer Sprachverarbeitung kann sehr konstruktiv sein.

NLP needs dictionaries, und dictionary-makers can use NLP to make better dictionaries, so there is a great potential for synergy between the two activities. [...] To date, NLP research has used dictionaries und dictionaries have used NLP, but the two processes have not been brought together. (Kilgariff 2000: 105)

Auf das genaue Zusammenwirken kann in diesem Zusammenhang nicht weiter eingegangen werden. Jedoch ist im Sinne einer konstruktiven Schnittstelle für das Modellierungskonzept festzuhalten, dass es hinreichend streng sein sollte, sodass die lexikografischen Daten gut automatisch weiterverarbeitet werden können und dass der inhaltliche Gehalt der Daten so maschinenlesbar transparent gemacht werden sollte, dass er für andere Anwendungen ausgenutzt werden kann.

Wie bereits ausgeführt, sollen die Daten ohne Strukturanzeiger, also ohne typografische Auszeichnungen etc. eingegeben werden können; gleichzeitig soll der inhaltliche Gehalt der Daten kodiert werden. Eine Trennung von Inhalt und Layout ist also eine der zentralen Anforderungen an die Modellierung.

2.2 Das Modellierungskonzept und seine Anwendung in *lexiko*

Die Modellierung erfolgt in *lexiko* XML-basiert. XML (eXtensible Markup Language) ist eine international festgelegte Syntax, eine Metasprache, die die

Entwicklung von Auszeichnungssprachen zur Beschreibung von Daten hinsichtlich ihrer hierarchischen Struktur und ihrer inhaltlichen Einheiten ermöglicht. (XML Standard [Third Edition]) Die Wahl von XML und verwandten Standards als Modellierungssprachen ist hier in verschiedener Hinsicht sinnvoll: Eines der wichtigsten Ziele von XML bzw. des Vorläufers SGML (Standard Generalized Markup Language) ist gerade die Trennung von Inhalt und Lay-out der Daten. Deshalb ist XML bzw. SGML schon früh für Verlage interessant gewesen:

The use of a single, internationally agreed syntax was not the only reason why a generic markup language was the appropriate solution für publishers that wanted to publish to multiple media. A fundamental concept in the use of generic markup languages is a move from formatting markup to content markup. (Gennusa 1999: 28)

Außerdem bietet XML als weltweit anerkannter Industriestandard eine bestmögliche Garantie dafür, dass die Daten langfristig weiterverwendet werden können. Eine XML-basierte Modellierung erfolgt in Form einer DTD (Document Type Definition) oder einem XML-Schema (das Für und Wider dieser beiden Formen ist hier nicht zu diskutieren). In einer solchen XML-DTD oder einem XML-Schema werden die Regeln festgelegt, der alle dazugehörigen Instanzen, d. h. in lexikografischen Projekten meist alle Artikel, zu gehorchen haben. Dies kann über Parsing-Prozeduren in XML-basierter Software geprüft werden. Neben vielen anderen Punkten, die für die Wahl von XML sprechen, wie z. B. die Verfügbarkeit von Software, ist daher v. a. wichtig, dass der Einsatz einer XML-basierten Modellierung eine gute Strukturführung der Lexikografen beim Erarbeiten der Artikel ermöglicht.

Allein mit dem Einsatz von XML ist das Modellierungskonzept jedoch noch kaum beschrieben, da XML auf viele Weisen eingesetzt werden kann. Daher muss genauer spezifiziert werden, wie die Modellierung erfolgen soll. Modellierungskonzepte unterscheiden sich von ihrer Intention her u. a. darin, ob sie allgemeine Richtlinien für eine Modellierung festlegen, die dann maßgeschneidert für ein Projekt umgesetzt werden, oder einen konkreten Modellierungsvorschlag entwickeln, der für möglichst viele Projekte anzuwenden ist, also eine maßgeschneiderte vs. eine Standard-Modellierung. Der bekannteste Vorschlag für eine Standard-Modellierung ist die Wörterbuchstruktur der TEI (Text Encoding Initiative), einen neuen Vorschlag dazu gibt Franziskus Geeb mit leXeML (Geeb 2001). Ein Vorteil dieser Initiativen oder Vorschläge besteht darin, dass keine eigene Modellierung im Projekt entwickelt werden muss. Nachteilig ist jedoch, dass entweder projektspezifisch starke Anpassungen vorgenommen werden müssen oder die Struktur sehr allgemein und – im Fall der TEI – sehr weich ist (s. u. a. Schmidt/Müller 2001: 37ff.). Eine solche Struktur bietet daher meist nicht in dem Maße eine Unterstützung im lexikografischen Prozess, wie das eine maßgeschneiderte Modellierung kann.

In *lexiko* wurde angestrebt, dass das Inhaltsstrukturenprogramm so genau wie möglich in der Modellierung abgebildet werden sollte, um den Lexikografen eine bestmögliche Strukturführung zu bieten. Ein Grund ist, dass voraussichtlich über einen längeren Zeitraum hinweg in verschiedenen personellen Zusammensetzungen an Wörterbuchartikeln gearbeitet wird, sodass eine möglichst umfangreiche automatische Konsistenzkontrolle der Daten sehr wichtig ist. Für *lexiko* wurde daher eine maßgeschneiderte Modellierung entwickelt. Erklärtes Ziel dieses Modellierungskonzeptes ist es, den inhaltlichen Gehalt der lexikografischen Daten und damit auch den genuinen Zweck (i. S. v. Wiegand 1989: 426), weshalb sie von den Lexikografen für die potenziellen Benutzer angesetzt wurden, so genau wie möglich zu kodieren und transparent zu machen. Unsere Hypothese ist dabei, dass potenziell jede angesetzte Angabe für einen gezielten Zugriff durch potenzielle Benutzer interessant sein kann. Dieser genuine Zweck liefert im Rahmen dieses Modellierungskonzeptes auch die Benennung für die jeweilige Klasse von Daten. Die Auszeichnung der lexikografischen Daten ist damit sehr granular, wodurch sie flexibel und gezielt zugreifbar werden. Die Modellierung ist darüber hinaus möglichst streng, d. h. in ihr wird so genau wie möglich festgelegt, welche Angaben in welcher Reihenfolge in den *lexiko*-Artikeln zu erarbeiten sind. Diese Reihenfolge bezieht sich allerdings nur auf die Erarbeitung der Daten und noch nicht auf ihre Präsentation. Diese strenge Modellierung bietet die geforderte Unterstützung im lexikografischen Prozess, da formal geprüft werden kann, ob die Artikelstruktur eingehalten wurde oder nicht.

2.3 Erarbeiten von Artikeln

Dies soll kurz demonstriert werden am Beispiel des Artikels *international*, der schon im ersten Teil des Aufsatzes erwähnt wurde. Abbildung 6 zeigt einen Screenshot des Artikels *international* im XMetaL-Editor, einem sehr weit verbreiteten Editor zum Erstellen von XML-Dateien. In einem solchen Editor wird beim Erstellen eines Artikels die zugehörige DTD geladen, in der die Regeln für die Artikelerstellung festgelegt sind.

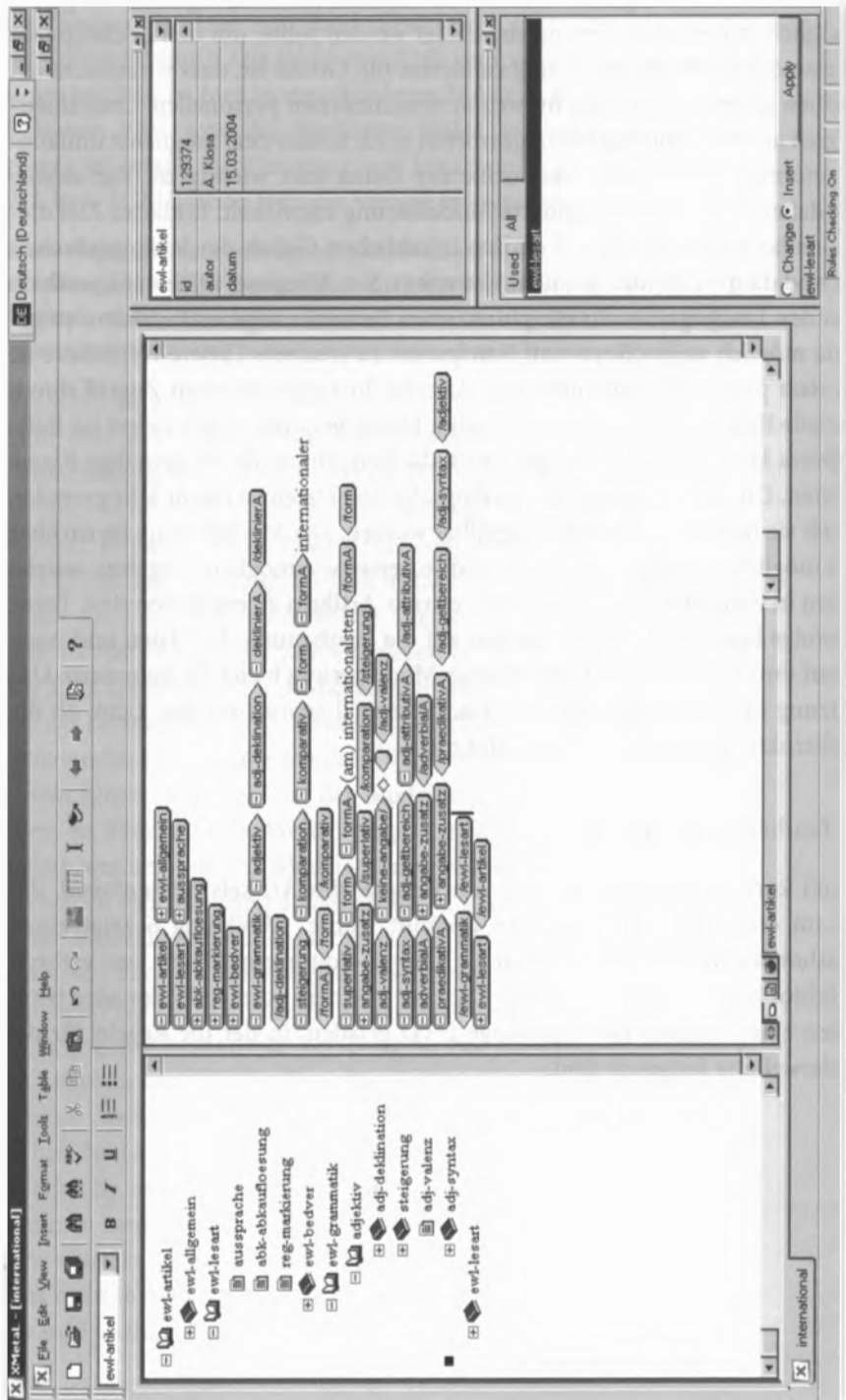


Abbildung 6. Bildschirmansicht des Artikels “international” im XMetal-Editor

Auf der linken Seite des Bildschirms sieht man einen Strukturüberblick über die Teile des Artikels. Es führt in diesem Zusammenhang zu weit, diese zu Grunde liegende DTD-Struktur detailliert zu erläutern. Für einen kurzen Überblick nur so viel: Es handelt sich hier um die Struktur für ein Einwortlemma, daneben gibt es in *lexiko* Mehrwortlemmata wie "toter Hund" und Wortelementlemmata wie "-lich" oder "-heit". Die Angaben zu einem ewl-artikel, d. h. einem Einwortlemma-Artikel, sind in zwei große Gruppen aufgeteilt: in lesartenübergreifende Angaben, die unter dem Element `ewl-allgemein` gefasst werden, und lesartenbezogene Angaben, die mit dem Element `ewl-lesart` ausgezeichnet werden. In *lexiko* wird auch die Grammatik lesartenbezogen angegeben; deshalb ist das Element `ewl-grammatik` unter den lesartenbezogenen Angaben angeordnet. Innerhalb von `ewl-grammatik` sieht man hier die einzelnen Angabegruppen, die zu Adjektiven gegeben werden und die zum Teil schon im ersten Teil dieses Artikels erläutert wurden: Angaben zur Deklinierbarkeit, zur Steigerung, zur Valenz und zur Syntax. In der Mitte sieht man den Artikel selbst, auch hier kann man Teile des Artikels ein- und ausblenden. Rechts oben sieht man Attribute, die Elementen zugeordnet werden, und rechts unten nach den Regeln der DTD auszuwählende Elemente.

Nun wurde schon mehrfach herausgestellt, dass die Lexikografen beim Eingeben der Artikel durch die Modellierung in dem Sinne unterstützt würden, dass die formale Einhaltung der Artikelstruktur geprüft wird. Dies funktioniert in der *lexiko*-Umgebung folgendermaßen: Die Lexikografen bekommen an der jeweiligen Stelle im Artikel, an dem sie gerade arbeiten, nur die Angaben im Editor angeboten, die durch die DTD an der jeweiligen Stelle erlaubt sind. Will ein Lexikograf beispielsweise die Grammatik zu *international* in der Lesart 'überstaatlich' im Artikel ausfüllen, muss er zunächst die Wortart bestimmen, hier also `adjektiv`. Abhängig von der Wortart öffnen sich dann jeweils unterschiedliche Angabeklassen. Bei Adjektiven sind es Angaben zur Deklinierbarkeit, zur Steigerung, zur Valenz und zur Syntax. Unter der `deklinierA` muss beispielsweise über ein Attribut ausgewählt werden, ob *international* in der Lesart 'überstaatlich' deklinierbar ist oder nicht. Als Demonstration nehmen wir an, dass der bearbeitende Lexikograf vergisst, den passenden Attributwert auszuwählen. Daraufhin muss er etwas zur Steigerung aussagen: entweder er kann `keine-angabe` geben oder er wählt das Element `aus`, falls das Adjektiv steigerbar ist. Wählt er das Element `komparation` aus, muss er den Komparativ und den Superlativ angeben. Zur Valenz gibt es im Fall von *international* `keine-angabe`. Unter Syntax werden die im ersten Teil beschriebenen Angaben zum Geltungsbereich gegeben. Man wählt also das Element `adj-geltbereich` aus und sieht rechts unten die Auswahl von `attributiv`, `adverbial` und `prädikativ`. Im Falle von *international* sind alle drei Möglichkeiten des Geltungsbereiches nacheinander auszuwählen. Mit dieser Auswahl sind die Angaben zum Geltungsbereich gemacht.

An diesem Beispiel kann man auch erkennen, dass in der Modellierung unterschieden wird zwischen absolut obligatorischen, relativ obligatorischen und fakultativen Angaben. Diese Unterscheidung geht zurück auf Wiegand (1989). Absolut obligatorische Angaben sind dabei solche Angaben, die immer gegeben werden sollen und auch gegeben werden können, wie z. B. eine Lemmazeichengestaltangabe, oder hier bei Adjektiven die Angabe zur Deklinierbarkeit. Relativ obligatorische Angaben sind dagegen Angaben, die nicht zu jedem Lemmazeichen gegeben werden können, wie z. B. Komplemente bei Adjektiven, die aber stets angegeben werden sollen, wenn es möglich ist. In der Modellierung werden diese relativ obligatorischen Angaben von rein fakultativen Angaben folgendermaßen unterschieden: Die Angabe selbst wird in der DTD, d. h. im 'Regelwerk' für die Erstellung der Artikel, – wie die absolut obligatorischen Angaben – als obligatorisch definiert. Innerhalb dieser Angabe oder Angabengruppe gibt es allerdings ein Ausweichelement *keine-angabe*. Damit wird deutlich gemacht, dass diese Angabe oder Angabengruppe immer bearbeitet werden soll und Lexikografen Stellung dazu nehmen müssen; anders als dies bei fakultativen Angaben der Fall ist. Mit dieser Modellierung soll auch gewährleistet werden, dass diese Angaben nicht versehentlich vergessen werden.

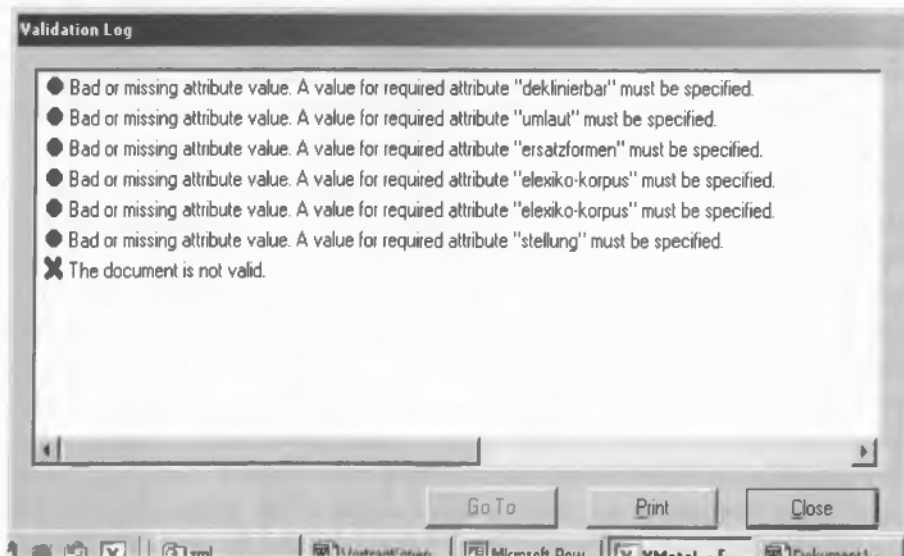


Abbildung 7. Fehlermeldung beim Validieren des Artikels

Ist der Lexikograf mit seinen Angaben zur Grammatik fertig und will den Artikel abschließen, muss er den Artikel zunächst validieren. Das bedeutet, dass der eingegebene Artikel gegenüber der in der DTD festgelegten Struktur geprüft wird. Da die Modellierung in *lexiko* sehr genau und streng ist, kann an dieser Stelle geprüft werden, ob der Lexikograf die festgelegte Artikelstruktur eingehalten hat oder nicht. Führt er diese Validierung im eben erläuterten Artikel durch, bekommt er die in Abbildung 7 gezeigten Fehlermeldungen.

Zunächst muss er also den Attributwert zur Angabe der Deklinierbarkeit ausfüllen, die er vorhin vergessen hatte. Dann wurde übersehen, dass zum Element *komparation* Attribute zu *umlaut* und *ersatzformen* auszufüllen sind. Diese sind im Falle von *international* beide auf "nein" zu setzen, weil die Steigerungsformen regelmäßig sind, also ohne Umlaut und Ersatzformen gebildet werden. Außerdem wurde bei den Steigerungsformen noch nicht vermerkt, ob sie im *lexiko*-Korpus belegt sind oder nicht. Dies muss aber in *lexiko* – wie vorhin ausgeführt – zu allen Formen gesagt werden. Die Steigerungsformen von *international* sind in unserem Korpus belegt, also werden die Attributwerte auf "ja" gesetzt. Die letzte Fehlermeldung besagt schließlich, dass für den Geltungsbereich *attributiv* etwas zur Stellung ausgesagt werden muss: für *international* ist dabei *prae nominal* auszuwählen. Validiert der Lexikograf den Artikel nun erneut, ist die Prüfung erfolgreich. So werden die Lexikografen unserer Erfahrung nach sehr gut in der formalen Einhaltung der Artikelstruktur unterstützt.

2.4 Flexibilität der Darstellung

Es wurde immer wieder herausgestellt, dass diese granulare Modellierung dazu dienen soll, die Daten der lexikografischen Datenbasis auf verschiedene Weise im elektronischen oder gedruckten Wörterbuch präsentieren zu können. Zur Darstellung der in XML vorliegenden lexikografischen Daten wird die Extensible Styling Language, kurz XSL, eingesetzt. Mit ihr lassen sich so genannte Stylesheets erstellen, in denen die Darstellung der Daten festgelegt wird. Im Prinzip kann dabei für jedes Element und jedes Attribut spezifiziert werden, wie es den potenziellen Benutzern dargestellt werden soll. Der Vorteil ist dabei, dass separat zu den eigentlichen Daten die Darstellung der Daten spezifiziert wird. Je granularer dabei die Datenauszeichnung ist, desto spezifischer kann demgemäß die Präsentation definiert werden. Außerdem können mehrere Stylesheets für die gleiche Datenbasis erstellt werden, so dass die gleiche Datenbasis in verschiedener Weise präsentiert werden kann. Zum Beispiel könnte ein Stylesheet für ein elektronisches Wörterbuch und eines für ein gedrucktes Wörterbuch erstellt werden. Darüber hinaus ist es möglich und auch sinnvoll, dass für ein elektronisches Wörterbuch mehrere Stylesheets erstellt werden, die adaptiv zu Benutzungssituationen eingesetzt werden.

Dieses Prinzip soll hier kurz an einem Beispiel erläutert werden, und zwar am Artikel *international* anhand der grammatischen Angaben in der Lesart 'überstaatlich'. Im Moment wird in *lexiko* nur ein Stylesheet verwendet, in dem die Darstellung aller Angaben im Artikel festgelegt wird. Die Darstellung der grammatischen Angaben im Artikel *international* in der Lesart 'überstaatlich' ist in Abbildung 2 gezeigt. Die gleichen Daten könnten aber auch durch eine Veränderung des Stylesheets in anderer Weise präsentiert werden. Für DaF-Nutzer könnte man z. B. festlegen, dass die Angaben wie in Abbildung 8 präsentiert werden. Hier wird die Angabe zur Steigerung verständlicher ausformuliert, und zu den Angaben zum Geltungsbereich sind die entsprechenden Belege standardmäßig geöffnet, um die Angaben besser zu verdeutlichen.

Für Experten hingegen könnte man ein weiteres Stylesheet einsetzen, in dem die Präsentation wie in Abbildung 9 gezeigt spezifiziert wird. Hier werden nur kurz die grundlegenden Eigenschaften in anderer Terminologie ("deklinabel, komparabel") genannt und die beobachteten syntaktischen Funktionen aufgezählt. Belege werden hier nicht mit angeboten. Diese Darstellungsmöglichkeiten sind für die genannten Benutzergruppen nicht zu Ende gedacht; sie sollen hier mehr der Demonstration dienen.

So können also für unterschiedliche Benutzergruppen in unterschiedlichen Benutzungssituationen unterschiedliche Darstellungen festgelegt werden, die alle auf die gleiche lexikografische Datenbasis zurückgreifen. Diese unterschiedlichen Darstellungen können ohne Veränderung der zu Grunde liegenden Datenbasis erreicht werden. Die Kombination von granular inhaltlich ausgezeichneten Daten und deren separat spezifizierte Darstellung ermöglicht daher eine sehr weit reichende Flexibilität der Darstellung. Denn so ist es möglich, abhängig von Benutzungssituationen und abhängig von bestimmten Suchanfragen nur eine Teilmenge von Daten anzeigen zu lassen, die Daten unter terminologisch anderen Überschriften zu präsentieren etc. Auch ist es möglich, für die Darstellung im gedruckten Wörterbuch die typografische Realisierung und das Generieren von Strukturanzeigern auf diese Weise festzulegen.

Internet Explorer - international - Microsoft Internet Explorer

Datei Bearbeiten Ansicht Favoriten Extras ?

Zurück - Suchen Favoriten Medien

Adresse: Eigene Dateien\Vorläge_Publikationen\Kopenhagen\Vortrag_Kopenhagen\Beispiele für Vortrag\Kopie von Flexibilität der Darstellung\international-lesarten.xml

Wechseln zu Links

Grammatik

Info

Adjektiv, das gesteigert werden kann:

internationaler
(am) *internationalsten*

Mögliche Verwendungen:

attributiv
adverbial

prädikativ

Beleg(e)

Beleg(e)

Wie von der Bundesregierung geplant und von Warschau nicht gewollt, verhandeln Kohl und Mazowiecki morgen parallel über den Grenzvertrag und die Rechte der Vertriebenen - beides wird gekoppelt, obwohl Bonn *international* zugesichert hat, die Grenze ohne Wenn und Aber anzuerkennen.
(die tageszeitung, 07.11.1990, S. 4, Polen-Vertrag. Bonn setzt sich durch.)

Schließen

Beleg(e)

Und da im Internet nicht nach Nationalitäten unterschieden wird, könnte aus dem Netz-Dauer-TED eine Nationalhymne entstehen, die wahrhaft *international* ist - so wie ja einst die Internationale die sowjetische Hymne gewesen ist, bevor der Zweite Weltkrieg dazwischen kam.
(Berliner Zeitung, 23.11.2000, Tagebuch, S. 13)

Schließen

Fertig Arbeitsplatz

Abbildung 8. Mögliche veränderte Darstellung der grammatischen Angaben im Artikel "international" für DaF-Nutzer

Lexikon Lesartenansicht: international - Microsoft Internet Explorer

Datei Bearbeiten Ansicht Favoriten Extras ?

Zurück - Suchen Favoriten Medien

Adresse: [Eigene Dateien\Vorträge_Publikationen\Kopenhagen\Vortrag_Kopenhagen\Beispiele für Vortrag\Flexibilität der Darstellung-Experten.xml|international-lesarten.xml] Wechselt zu Links

international Lesart 'überstaatlich'

Lesarten/überstättende Angaben
 Lesart 'überstaatlich'
 Lesart 'indischstaatlich'

Bedeutungs- erläuterung	Semantische Umgebung u. lexikalische Mitspieler	Typische Verwendungen	Sinnenwandelte Wörter	Besonderheiten des Gebrauchs	Grammatik
Grammatik <input type="button" value="info"/>					
Wortart:	Adjektiv				
Eigenschaften:	deklinabel komparabel				
Syntaktische Funktionen:	Attribut Adverbiale Prädikativkomplement				

Fertig

Abbildung 9. Mögliche veränderte Darstellung der grammatischen Angaben im Artikel "international" für Experten

2.5 Was noch zu tun ist

Auch in diesem Bereich des Projektes gibt es aber natürlich noch viel zu tun. Anfangs wurden die Anforderungen an ein Modellierungskonzept vorgestellt, die für die Entwicklung der Modellierung relevant waren. Mit dieser jetzt entwickelten und angewandten Modellierung ist auf Ebene der lexikografischen Datenbasis die Voraussetzung für einen sehr flexiblen Umgang mit den lexikografischen Daten auf Ebene der Präsentation geschaffen. Allerdings werden diese Möglichkeiten noch nicht in vollem Umfang in wirkliche Funktionalität für die Wörterbuchbenutzer umgesetzt.

Was jetzt schon als Vorteil dieser Modellierung zum Tragen kommt ist die – auch immer wieder geforderte – Unterstützung der Lexikografen bei der Einhaltung der formalen Artikelstruktur. Dies hat sich auch u. a. bei der Einarbeitung einer neuen Kollegin sehr gut bewährt. Doch das volle Potenzial der gezeigten Modellierung kann erst dann wirklich gezeigt werden, wenn man

1. adaptiv zu Benutzungssituationen flexible Darstellungsmöglichkeiten entwickelt und
2. sehr umfangreiche Recherchemöglichkeiten anbietet.

Recherchemöglichkeiten sind zwar ansatzweise angeboten, doch sie sind bei weitem noch nicht so umfassend, wie sie von der Aufbereitung der Daten her sein könnten. Das Problem ist also nicht (wie sonst oft in lexikografischen Projekten), dass die Daten nicht richtig aufbereitet sind, und von daher solche Möglichkeiten überhaupt nicht zur Verfügung stehen. Vielmehr ist auf dem Weg von der lexikografischen Datenbasis hin zur Präsentation der Daten im elektronischen Wörterbuch noch vieles zu tun.

Literatur

- Bergenholtz, Henning und Joachim Mugdan. 1984a. "Vorwort". In *Studien zur neuhochdeutschen Lexikographie V*, H. E. Wiegand (Hg.) [Germanistische Linguistik 3–6, 1984], 7–19. Hildesheim, Zürich und New York: Georg Olms.
- Bergenholtz, Henning und Joachim Mugdan. 1984b. "Grammatik im Wörterbuch: Von Ja bis Jux". In *Studien zur neuhochdeutschen Lexikographie V*, H. E. Wiegand (Hg.) [Germanistische Linguistik 3–6, 1984], 47–75. Hildesheim, Zürich und New York: Georg Olms.
- Büchel, Gregor und Bernhard Schröder. 2001. "Verfahren und Techniken in der computergestützten Lexikographie". In *Chancen und Perspektiven computergestützter Lexikographie*, I. Lemberg/B. Schröder/A. Storrer (Hgg.) [Lexicographica. Series Maior 107], 7–28. Tübingen: Niemeyer.
- Farø, Ken. 2005. "Korpuslexikographie und Flexionsmorphologie". In *Symposium on Lexicography XI. Proceedings of the Eleventh International Symposium on Lexicography*. May, 2–4,

- 2002 at the University of Copenhagen, H. Gottlieb/J. E. Mogensen/A. Zettersten (Hgg.) [Lexicographica. Series Maior 115], 217–226. Tübingen: Niemeyer.
- Geeb, Franziskus. 2001. "leXeML – Vorschlag und Diskussion einer (meta-)lexikographischen Auszeichnungssprache". *Sprache und Datenverarbeitung* 2 (2001): 27–61.
- Gennusa, Pamela L. 1999. "Evolution and use of generic markup languages". In *SGML und XML. Anwendungen und Perspektiven*, W. Möhr/I. Schmidt (Hgg.), 27–50. Berlin/Heidelberg/New York: Springer.
- Gloning, Thomas und Rüdiger Welter. 2001. "Wortschatzarchitektur und elektronische Wörterbücher: Goethes Wortschatz und das Goethe Wörterbuch". In *Chancen und Perspektiven computergestützter Lexikographie*, I. Lemberg/B. Schröder/A. Storrer (Hgg.) [Lexicographica. Series Maior 107], 117–132. Tübingen: Niemeyer.
- Heyn, Matthias. 1992. *Zur Wiederverwendung maschinenlesbarer Wörterbücher. Eine computergestützte metalexikographische Studie am Beispiel der elektronischen Edition des "Oxford Advanced Learner's Dictionary of Current English"* [Lexicographica. Series Maior 45]. Tübingen: Niemeyer.
- Kilgariff, Adam. 2000. "Business Models for Dictionaries und NLP". *International Journal of Lexicography* 13 (2/2000): 107–118.
- Klosa, Annette. 2001. "Qualitätskriterien der CD-ROM-Publikation von Wörterbüchern". In *Chancen und Perspektiven computergestützter Lexikographie*, I. Lemberg/B. Schröder/A. Storrer (Hgg.) [Lexicographica. Series Maior 107], 93–101. Tübingen: Niemeyer.
- Mugdan, Joachim. 1989. "Information on Inflectional Morphology in the General Monolingual Dictionary". In *Wörterbücher. Ein internationales Handbuch zur Lexikographie*. 1. Teilband [Handbücher zur Sprach- und Kommunikationswissenschaft 5.1], F. J. Hausmann/O. Reichmann/H. E. Wiegand/L. Zgusta (Hgg.), 518–525. Berlin und New York: De Gruyter.
- Schmidt, Ingrid und Carolin Müller. 2001. "Entwicklung eines lexikographischen Modells: Ein neuer Ansatz". In *Chancen und Perspektiven computergestützter Lexikographie*, I. Lemberg/B. Schröder/A. Storrer (Hgg.) [Lexicographica. Series Maior 107], 29–52. Tübingen: Niemeyer.
- Schryver, Gilles-Maurice de. 2003. "Lexicographer's Dreams in the Electronic-Dictionary Age". *International Journal of Lexicography* 16 (2/2003): 143–199.
- Storrer, Angelika. 2001. "Digitale Wörterbücher als Hypertexte: Zur Nutzung des Hypertextkonzepts in der Lexikographie". In *Chancen und Perspektiven computergestützter Lexikographie*, I. Lemberg/B. Schröder/A. Storrer (Hgg.) [Lexicographica. Series Maior 107], 53–69. Tübingen: Niemeyer.
- Trap-Jensen, Lars. 2002. "Descriptive and Normative Aspects of Lexicographic Decision-Making: The Borderline Cases". In *Proceedings of the Tenth EURALEX International Congress. Vol. II*, A. Braasch/C. Povlsen (Hgg.), 503–508. Kopenhagen: CST.
- Wiegand, Herbert Ernst. 1985. "Fragen zur Grammatik in Wörterbuchbenutzungsprotokollen. Ein Beitrag zur empirischen Erforschung der Benutzer einsprachiger Wörterbücher". In *Lexikographie und Grammatik, Akten des Essener Kolloquiums zur Grammatik im Wörterbuch 28.-30.6.1984*, H. Bergenholz/J. Mugdan (Hgg.) [Lexicographica. Series Maior 3], 20–98. Tübingen: Niemeyer.
- Wiegand, Herbert Ernst. 1989. "Der Begriff der Mikrostruktur: Geschichte, Probleme, Perspektiven". In *Wörterbücher. Ein internationales Handbuch zur Lexikographie*. 1. Teilband [Handbücher zur Sprach- und Kommunikationswissenschaft 5.1], F. J. Hausmann/O. Reichmann/H. E. Wiegand/L. Zgusta (Hgg.), 409–462. Berlin und New York: De Gruyter.

Wiegand, Herbert Ernst. 1998. *Wörterbuchforschung. Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie*. 1. Teilband. Berlin und New York: De Gruyter.

XML Standard [Fourth Edition]: www.w3.org/TR/REC-xml. (27.7.2007)

Zifonun, Gisela; Ludger Hoffman und Bruno Strecker. 1997. *Grammatik der deutschen Sprache*. Berlin und New York: De Gruyter.